Modaalsed loogikad

Modaalsed loogikad: idee

- Harilikus loogikas esitavad laused mingeid väiteid
- Samas on keeruline (või võimatu) öelda samas loogikas midagi nende lausete ENDI kohta:
  - kui tõenäoliselt kehtib
  - kindlasti/võibolla kehtib
  - kellegi arvates kehtib
  - minevikus/tulevikus kehtib

- Modaalsetes loogikates saab öelda “metaväiteid”, so teatud väiteid teiste väidete kohta, ehk asju nagu:
  - On võimalik, et “A=>B või C”
  - Tulevikus kehtib igal juhul, et “A & B => C”
  - A usub, et “B & C”
  - Jne

Modaalsed loogikad: palju eri tüüpe

- Kuna võiksime soovida öelda erinevat tüüpi metaväiteid (eelmine slaid), siis on ka palju erinevaid modaalseid loogikaid:
  - **Modal logic**
    - [] A It is necessary that A
    - ◊ A It is possible that A
  - **Deontic Logic**
    - O A It is obligatory that A
    - P A It is permitted that A
    - F A It is forbidden that A
  - **Temporal Logic**
    - G A It will always be the case that ..
    - F A It will be the case that ..
    - H A It has always been the case that ..
    - P A It was the case that..
  - **Doxastic Logic**
    - B x A x believes that A

“Harilikud” modaalsed loogikad: perekond loogikaid

- Kaks operaatorit:
  - [] A : **on paratamatult** A
  - ◊ A : **on võimalik**, et A
- NB! [] A = ¬ ◊ ¬ A ja □ A = ¬ [] ¬ A

- Näiteks:
  - **“Dogs are dogs”** on paratamatult õige.
  - **“Dogs are pets”** on õige, aga mitte paratamatult õige

- Mida aga **täpselt** tähendab “**on paramatult**” ja “**on võimalik**”??
- Palju erinevaid võimalikke intepretatsioone!
- Seepärast ka hulk erinevaid modaalseid loogikaid.

“Harilike” modaalsete loogikate kihiline ehitus

- Kõige nõrgem: süsteem **K**. Sisaldab:
  - **Klassikalist loogikat**
  - **Järeldusreeglit: If A is a theorem of K, then so is [] A**
    - Näiteid/seletusi selle järeldusreegli kohta:
      - (A => A) on lausearvutuse teoreem, seega ka K teoreem, seega järeldame reegliga, et [] (A => A). Siit saame samm-sammult edasi järeldada []... [] (A => A)
      - Samas, “lihtsalt“ A ei ole teoreem. Seega me ei saa järeldada [] A.
      - Kas saame järeldada A => [] A ? Ei saa! Ainult juhu jaoks, kui A on tautoloogia, st ise järeldub teooriast.
  - **Aksioomi:** **[] (A => B) => ([] A => [] B).**

“Harilike” modaalsete loogikate kihiline ehitus

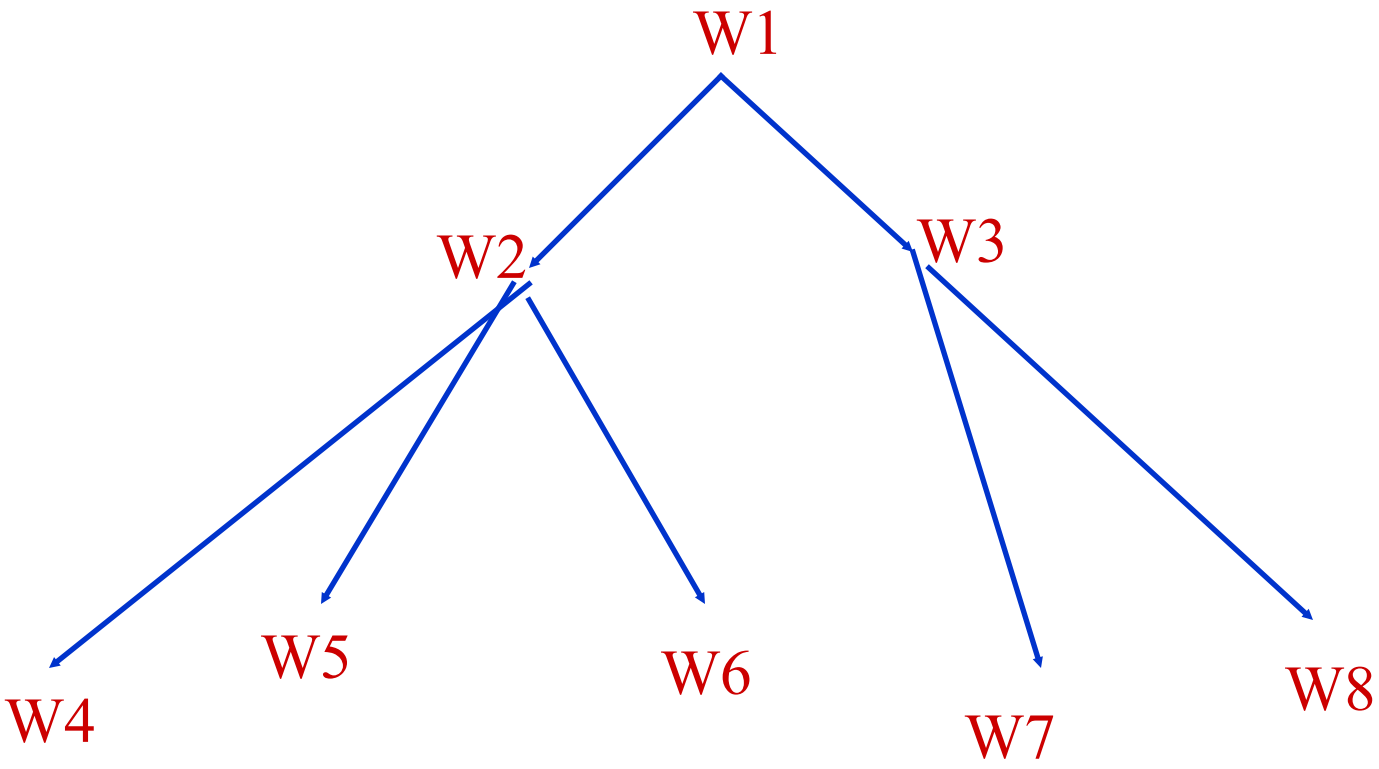
- Kõige nõrgem: süsteem **K**. Sisaldab:
  - Klassikalist loogikat
  - Järeldusreeglit: If A is a theorem of K, then so is [] A
  - Aksioomi: [] (A => B) => ([] A => [] B).

- Järgmine: süsteem **M**. Sisaldab LISAKS K-le:
  - [] A => A

- Järgmine, alternatiiv: **S4**. Sisaldab lisaks M-le:
  - [] A => [] [] A (seega alati [] ...[] A = [] A, sama ka ◊ jaoks)
- Järgmine, alternatiiv: **S5**. Sisaldab lisaks M-le:
  - □ A => [] □ A
    - (seega alati [] ◊ ..□ .[] .. ◊ A = □ A, ja [] □ ..□ .[] .. □ [] A = [] A)
- S5 saab S4-st, kui lisada A => [] □ A

Possible worlds semantics (originating from Kripke)

- Põhiidee: võimalike tulevike/olukordade/maailmade puu:



- Harilikud lausearvutuse tõeväärtustabelid, aga iga maailma (Wi) jaoks eraldi!!

Possible worlds semantics: tõeväärtustabel I

- Olgu meil lausemuutujad A ja B. Võimalik tõeväärtustabel nende jaoks oleks siis näiteks:

World	A	B
W1	T	F
W2	T	T
W3	F	F
W4	T	T
W5	T	F
W6	T	T
W7	F	F
W8	T	T

Possible worlds semantics: valemi tõeväärtus

- Olgu meil mingi suurem valem. Mis on selle tõeväärtus mingi tõeväärtustabeli jaoks?
- Tähistame **v(p,w):** valemi p tõeväärtus maailmas w.
- Siis:

(¬) v(¬ **A**, w)=T iff v(**A**, w)=F.

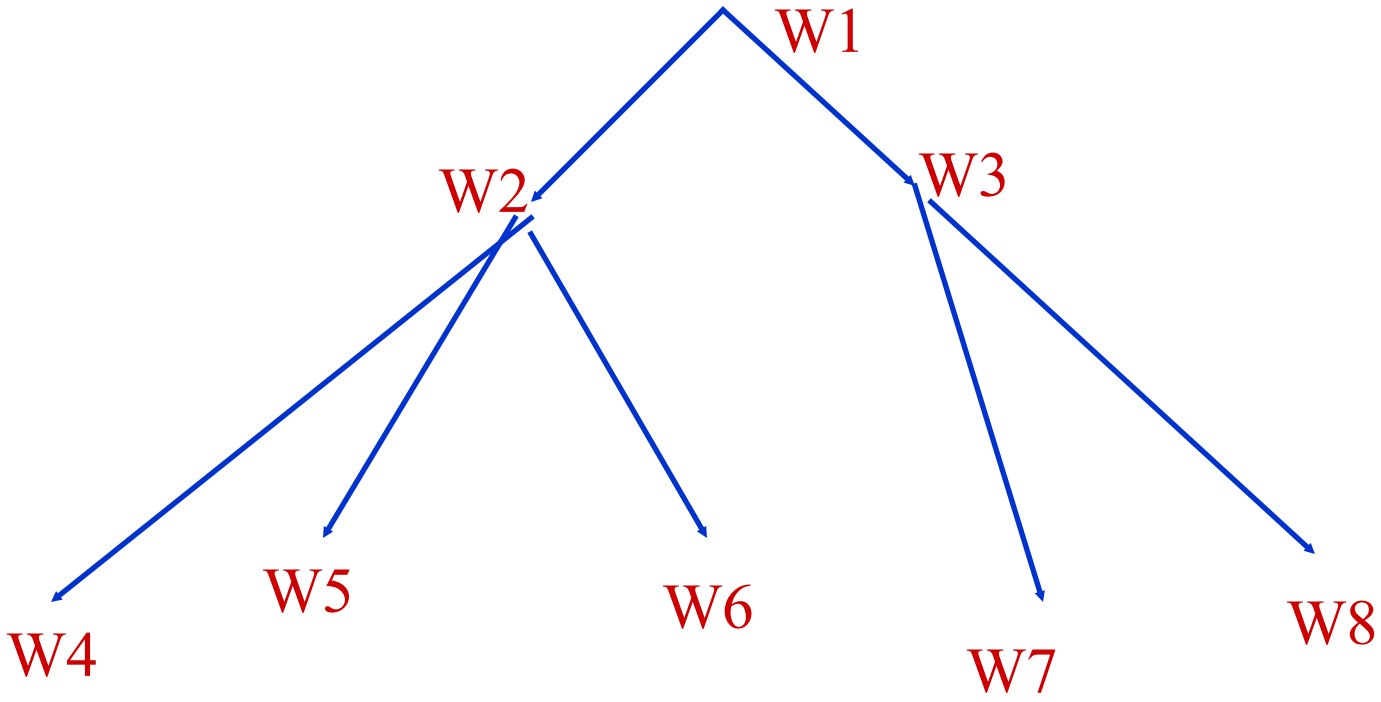
(=>) v(**A=>B**, w)=T iff v(**A**, w)=F or v(**B**, w)=T.

(**5**) v(□ A, w)=T iff for every world w in W, v(**A**, w)=T.

- An argument is said to **5**-*valid* iff it is valid for every non empty set of W of possible worlds.
- It has been shown that **S5** is sound and complete for **5**-validity

Possible worlds semantics: S5 probleem

- S5 ei arvesta võimalikke piiranguid maailmast maailma minekul: S5 jaoks on tegu lihtsalt maailmade hulga, mitte orienteeritud puuga.



- Oletame, et tahame rääkida ajast. Siis on meil tegu maailmade puuga, mitte lihtsalt maailmade hulgaga.

Semantics: alternatiiv reeglile 5

- Kui tahame piirata maailmast maailma minekut, sobib 5 asemel semantikasse reegel K:

(K) v([] **A**, w)=T iff for every w' , if w**R**w' , then v(**A**, w')=T.

wRw' tähendab: maailmast w saab minna maailma w'

- This says that A is true at w just in case A is true at all times *after* w.
- Tõeväärtustabelile tuleb lisada maailmast-maailma mineku tabel R.
- Loogika K on (K) reegli jaoks korrektne ja täielik.

Semantics: täiendatud alternatiiv reeglile 5

- Vaatasime: kui tahame piirata maailmast maailma minekut, sobib 5 asemel semantikasse reegel K:

(K) v([] **A**, w)=T iff for every w' , if w**R**w' , then v(**A**, w')=T.

wRw' tähendab: maailmast w saab minna maailma w'

- Tõeväärtustabelile tuleb lisada maailmast-maailma mineku ruudukujuline tabel R.
- Kui me nüüd aga NÕUAME, et **R oleks transitiivne** (kui A-st saab B-sse ja B-st saab C-sse, siis A-st saab C-sse), siis klapib selle semantikaga loogika S4.
- NB! Transitiivsus kehtib näiteks ajaloogikate puhul.

Teadmiste läbipaistmatus I

- Suppose one wishes to reason about intentional notions in a logical framework. Consider the following statement (after [Genesereth and Nilsson, 1987]):

Janine believes Cronos is the father of Zeus (1)

- A naive attempt to translate (1) into first-order logic might result in the following:

Bel(Janine, Father(Zeus,Cronos)) (2)

- Unfortunately, this naive translation does not work, for two reasons.
  - **The first is syntactic**: the second argument to the Bel predicate is a *formula* of first-order logic, and is not, therefore, a term. So (2) is not a well-formed formula of classical first-order logic.

Teadmiste läbipaistmatus II

- The second problem is semantic, and is potentially more serious. The constants **Zeus** and **Jupiter**, by any reasonable interpretation, denote the same individual: the supreme deity of the classical world. It is therefore acceptable to write, in first-order logic:

(Zeus = Jupiter) **(3)**

- Given (2) and (3), the standard rules of first-order logic would allow the derivation of the following:

Bel(Janine, Father(Jupiter,Cronos))

- But intuition rejects this derivation as invalid: **believing that the father of Zeus is Cronos is *not* the same as believing that the father of Jupiter is Cronos**. So what is the problem? Why does first-order logic fail here? The problem is that the intentional notions - such as belief and desire - are ***referentially opaque***, in that they set up ***opaque contexts***, in which the standard substitution rules of first-order logic do not apply,

Lahendusvariante

- The first, best-known, and probably most widely used approach is to adopt a ***possible worlds* semantics**, where an agent's beliefs, knowledge, goals, and so on, are characterized as a set of so-called *possible worlds*, with an *accessibility relation* holding between them. Possible worlds semantics have an associated *correspondence theory* which makes them an attractive mathematical tool to work with [Chellas, 1980]. However, they also have many associated difficulties, notably the well-known *logical omniscience* problem, which implies that agents are perfect reasoners.
- The commonest alternative to the possible worlds model for belief is to use a ***sentential*, or *interpreted symbolic structures*** approach. In this scheme, beliefs are viewed as symbolic formulae explicitly represented in a data structure associated with an agent. An agent then believes P if P is present in its belief data structure. Despite its simplicity, the sentential model works well under certain circumstances [Konolige, 1986a].

Variant I: Teadmised ja modaalsused

- Tradmised on sarnased modaalsustele.
  - [] **A** tähistab: **isik teab, et A on õige/kehtib**
- R ehk maailmast maailma minek vastab järelduste tegemise võimele.

Variant I: teadmised ja possible worlds: Hintikka

- **Agent's beliefs could be characterized as a set of *possible worlds*, in the following way.**
- **Consider an agent playing a card game such as poker** . In this game, the more one knows about the cards possessed by one's opponents, the better one is able to play. And yet complete knowledge of an opponent's cards is generally impossible, (if one excludes cheating). The ability to play poker well thus depends, at least in part, on the ability to deduce what cards are held by an opponent, given the limited information available.

Variant I: teadmised ja possible worlds: Hintikka

- **Now suppose our agent possessed the ace of spades**. Assuming the agent's sensory equipment was functioning normally, it would be rational of her to believe that she possessed this card. Now suppose she were to try to deduce what cards were held by her opponents. This could be done by first calculating all the various different ways that the cards in the pack could possibly have been distributed among the various players. (This is not being proposed as an actual card playing strategy, but for illustration!)
- **For argument's sake, suppose that each possible configuration is described on a separate piece of paper**. Once the process was complete, our agent can then begin to systematically eliminate from this large pile of paper all those configurations which are *not possible, given what she knows*.
- **For example, any configuration in which she did not possess the ace of spades could be rejected immediately as impossible. Call each piece of paper remaining after this process a *world*.**

Omniscience ehk ideaalne arutleja

- **Rule “If A is a theorem of K, then so is [] A“ tells us that an agent knows all valid formulae.**
- **Amongst other things, this means an agent knows all propositional tautologies**. Since there are an infinite number of these, an agent will have an infinite number of items of knowledge: immediately, one is faced with a counter-intuitive property of the knowledge operator.

Omniscience ehk ideaalne arutleja

- **Now consider the axiom K ([] (A => B) => ([] A => [] B))** , which says that an agent's knowledge is closed under implication. Together with the necessitation rule, this axiom implies that an agent's knowledge is closed under logical consequence: an agent believes all the logical consequences of its beliefs. This also seems counter intuitive. For example, suppose, like every good logician, our agent knows Peano's axioms. Now Fermat's last theorem follows from Peano's axioms - but it took the combined efforts of some of the best minds over the past century to prove it. Yet if our agent's beliefs are closed under logical consequence, then our agent must know it. So consequential closure, implied by necessitation and the K axiom, seems an overstrong property for resource bounded reasoners.
- **These two problems - that of knowing all valid formulae, and that of knowledge/belief being closed under logical consequence - together constitute the famous *logical omniscience* problem**. It has been widely argued that this problem makes the possible worlds model unsuitable for representing resource bounded believers - and any real system is resource bounded.

Teadmised: mitu isikut

- **Lisaparameeter [] -operatorile: isik.**
  - **teab(i,v)** tähistab, et “isik i aktsepteerib alati valemit v ja v on õige”
- Teadmised vs uskumised
  - **usub(i,v)** tähistab, et “isik i aktsepteerib alati valemit v”
  - **teab(i,v)** tähistab, et “isik i aktsepteerib alati valemit v ja v on õige”

Cohen and Levesque – intention I

- Following Bratman, [Bratman, 1990][Bratman, 1987], Cohen and Levesque identify seven properties that must be satisfied by a reasonable theory of intention:

1. Intentions pose problems for agents, who need to determine ways of achieving them.
2. Intentions provide a `filter' for adopting other intentions, which must not conflict.
3. Agents track the success of their intentions, and are inclined to try again if their attempts fail.
4. Agents believe their intentions are possible.
5. Agents do not believe they will not bring about their intentions.
6. Under certain circumstances, agents believe they will bring about their intentions.
7. Agents need not intend all the expected side effects of their intentions.

Cohen and Levesque – intention II

- Given these criteria, Cohen and Levesque adopt a two-tiered approach to the problem of formalizing intention. First, they construct a *logic of rational agency*, `being careful to sort out the relationships among the basic modal operators' [Cohen and Levesque, 1990a]. Over this framework, they introduce a number of derived constructs, which constitute a `partial theory of rational action' [Cohen and Levesque, 1990a]; intention is one of these constructs. The first major derived construct is the *persistent goal*. An agent has a persistent goal of P iff:

1. It has a goal that P eventually becomes true, and believes that P is not currently true.
2. Before it drops the goal P , one of the following conditions must hold: (i) the agent believes P has been satisfied; or (ii) the agent believes P will never be satisfied.

Võimalik arendussuund: tähenduse (puudulik) formaliseerimiseks

- Reaalne olukord tähenduse osas sisaldab lisaks muudele tüüpiliselt järmisi parameetreid:
  - **Isik**, kelle jaoks me tähendust rehkendame
  - **Aeg**, millal me tähendust rehkendame
  - **Tüüpkontekst**, mille suhtes me tähendust rehkendame
  - **Valem**, mille tähendust me rehkendame

- Valemid on nii tõeväärtusfunktsioonid kui “harilikud” funktsioonid (näiteks pluss, isa jne)

- Tähenduse funktsioon f(x1,...xn) võtab seega lisaparameetrid:

f(isik, aeg, kontekst, x1,...,xn)

Võimalik arendussuund: tähenduse (puudulikuks) formaliseerimiseks

- Kommunikatsioon (üks isik x ütleb teisele isikule y teate p) vajab lisaks asju:
  - **x teadmised y kohta**
  - **y teadmised x kohta**
- y mõistab x antud p-d tähendusfunktsiooni f abil:

p tähendus arvutatakse y poolt funktsiooni/programmiga:

f(x,y,aeg,teadmised(x,y),kontekst,p)

kus:

- **teadmised(x,y)** on need teadmised, mis x-i teada y-l on
- **p** on x-i poolt y-le antud teade